

Санкт-Петербургский государственный университет

Математическое обеспечение и администрирование информационных  
систем

Технология программирования

Дудкин Евгений Владимирович

# Интер- и интрасетевые связи: визуализация и анализ

Выпускная квалификационная работа

Научный руководитель:  
профессор кафедры информатики, д. ф.-м. н., доц. Тулупьев А. Л.

Рецензент:  
с.н.с., ТиМПИ СПИИРАН, к.ф.-м.н. Суворова А. В.

Санкт-Петербург  
2017

SAINT-PETERSBURG STATE UNIVERSITY

Software and Administration of Information Systems  
Technology of Programming

Evgeny Dudkin

# Inter- and intraconnections in networks: visualisation and analysis

Graduation Thesis

Scientific supervisor:  
Prof. Computer Science Department, Dc. Sc. in Math, Assoc. Prof. Alexander Tulupyeu

Reviewer:  
Senior Researcher, TICS Lab, SPIIRAS, PhD in Math Alena Suvorova

Saint-Petersburg  
2017

# Оглавление

<b>Введение</b>	<b>4</b>
<b>1. Основные понятия</b>	<b>5</b>
<b>2. Обзор существующих решений</b>	<b>6</b>
<b>3. Технологии</b>	<b>7</b>
<b>4. Описание решения</b>	<b>8</b>
4.1. Регистрация приложения . . . . .	8
4.2. Авторизация пользователя . . . . .	8
4.3. Поиск онлайн-сообществ . . . . .	10
4.4. Построение социальной сетевой схемы сообщества . . . .	12
4.5. Построение социальной сетевой схемы набора сообществ	15
4.6. Визуализация . . . . .	16
4.7. Создание CSV файла . . . . .	23
4.8. Анализ графа . . . . .	25
4.9. Пример использования . . . . .	26
<b>Заключение</b>	<b>36</b>
<b>Список литературы</b>	<b>38</b>

# Введение

В наши дни социальные сети стали неотъемлемой частью нашей жизни. Они являются источником данных и средством распространения информации. Поэтому в последнее время увеличилось количество исследований, проводимых в социальных сетях. При этом огромное внимание уделяется изучению онлайн-сообществ как особых социальных общностей [4], включая, например, вопросы, можно ли считать формальные онлайн-группы сообществами [8], исследования типов онлайн-сообществ [9], и другие исследования, основанные на особенностях структуры сообществ [16]. Также изучение сообществ позволяет выявлять как активных пользователей, вносящих существенный вклад в деятельность группы [5], так и «молчащих» пользователей, регулярно читающих публикуемый контент, но не участвующих в обсуждениях [10]. Причем именно вторая группа пользователей часто интересует исследователей, как с точки зрения изучения механизмов мотивации [2], так и с точки зрения потенциального рекрутирования в проводимые исследования [7].

Для проведения исследований онлайн-сообществ необходимы программы способные собирать необходимую информацию о группах. Но зачастую исследователям бывает недостаточно существующих инструментов для решения поставленных задач. Специалисты, составившие план изучения сообществ в социальных сетях, вынуждены тратить время и средства на сбор и анализ данных. Поэтому возникла необходимость в программном обеспечении, которое позволит собирать необходимые данные онлайн-сообществ, а также предоставит набор инструментов для их базового анализа.

## Постановка задачи

Цель данной работы — автоматизировать сбор, обработку и визуализацию информации об онлайн-сообществах социальной сети «ВКонтакте» (в рамках поставленных ниже задач).

Для достижения цели были поставлены следующие задачи:

- обеспечить сбор данных о структурах онлайн-сообществ социальной сети «ВКонтакте», включая список контактов, город, ссылки, страну и информацию об участниках;
- визуализировать структуру сообщества в виде графа, вершинами которого являются участники этого сообщества, а ребрами — соответствующие связи между этими пользователями (дружба, обучение в одном вузе, город, родной город, обучение в одной школе, наличие общих сообществ);
- визуализировать структуру связей между несколькими сообществами в виде графа, вершинами которого являются эти сообщества, а ребрами — соответствующие связи между ними (город, ссылки на одни и те же ресурсы или на друг друга, страна, наличие общих пользователей в графе контакты);
- предоставить инструменты для базового анализа полученных графов (вычисление показателей центральности, среднего для показателей центральности, количества вершин и ребер, вычисление количества изолированных вершин, количества компонент связности, количества вершин максимальной компоненты связности и др.) ;
- предоставить возможность импортировать данные о структурах в виде CSV файлов.

## 1. Основные понятия

Для анализа онлайн-сообществ, как социальных сетей, используются специальные метрики. Введем некоторые базовые понятия, связанные с анализом социальных сетей.

Центральность - метрика, целью которой является определение «значительности» или «влияния» (в различных значениях) определенного узла (или группы) в сети [11]. Самыми популярными методами для

измерения центральности являются: определение центральности по посредничеству (betweenness centrality), центральность по близости (closeness centrality), центральности собственного вектора (eigenvector centrality) и центральности по степени (degree centrality). Центральность узла по посредничеству равна числу кратчайших путей из всех вершин ко всем остальным, которые проходят через этот узел [11]. Центральность по близости рассчитывается как сумма длин кратчайших путей между узлом и всеми другими вершинами в сети [11]. Центральность собственного вектора измеряет, насколько объект связан с другими объектами и насколько он может непосредственно повлиять на другие соединенные объекты в сети [11]. Центральность по степени измеряет, насколько объект связан с другими объектами, подсчитывая число прямых связей каждого объекта с другими объектами в сети [11].

Взаимность - степень, с которой двое участников отвечают друг другу взаимностью в сфере дружеских или других взаимодействий [11].

Транзитивность измеряет вероятность того, что соседние вершины соединены. Это иногда также называют коэффициентом кластеризации [11].

## 2. Обзор существующих решений

В последнее время проводится все больше исследований данных, которые были получены с помощью социальной сети «ВКонтакте». Например, исследование 2016 года «Структура и функции онлайн-сообществ: сетевая картография ВИЧ-релевантных групп в социальной сети «ВКонтакте» » [17]. Или исследование 2013 года «Технологии политической мобилизации в социальной сети «ВКонтакте»: сетевой анализ протестного и провластного сегментов» [18]. Происходит это потому, что данная сеть является наиболее популярной социальной сетью в России [12]. Но уже на этапе сбора данных исследователи и научные деятели сталкиваются с рядом трудностей.

Во-первых, существующие средства для автоматизированного сбора данных чаще всего ориентированы на маркетинговые цели. В боль-

шинстве случаев они направлены на отслеживание упоминания бренда в сети или на предоставление некоторой статистики. Например, продукт YouScan [13] мониторит упоминания бренда товара в социальных сетях, блогах, форумах, сайтах для отзывов, а также в онлайн-СМИ. Или сервис мониторинга социальных медиа и онлайн-СМИ IQBuzz [3]. По словам разработчиков данный сервис дает возможность определять релевантные площадки, где наиболее часто ведется обсуждение компании, вычислить лидеров мнений и организовать взаимодействие с ними, отслеживать эффективность PR-кампаний и реакцию пользователей на информационный повод, анализировать деятельность конкурентов и отслеживать вброс негатива с их стороны [3]. Кроме того, инструменты сбора данных чаще всего являются коммерческими. Для полноценного использования IQBuzz и YouScan пользователь должен потратить определенную сумму денег. Как следствие, исследователям онлайн-сообществ, которым необходимо получать не просто данные о конкретном сообществе, а исследовать его структуру и связи, такие средства автоматизированного сбора данных не подходят.

Во-вторых, другие, более гибкие, инструменты сбора данных из социальной сети «ВКонтакте» (например, средства API) требуют некоторых технических навыков и навыков программирования. Но зачастую специалисты социогуманитарного цикла такими навыками не обладают.

### **3. Технологии**

Для решения поставленных задач понадобилось использовать следующие технологии: VK Api (методы для работы с данными «ВКонтакте») [15], Igraph (библиотека для анализа социальных сетей) [14], R.Net (библиотека для интеграции языка R в фреймворк .NET) [6]. Язык R был выбран потому, что он обеспечивает большое количество статистических и численных методов и имеет возможность расширения функционала с помощью пакетов. Пакеты — библиотеки для работы специфических функций. Также R обладает широкими графическими

возможностями для создания качественной графики. Это необходимо для задачи визуализации графов. Igraph — это набор инструментов сетевого анализа с акцентом на эффективность, мобильность и простоту использования. Igraph бесплатный проект с открытым исходным кодом. Кроме того, Igraph может быть запрограммирован в языке R.

## **4. Описание решения**

### **4.1. Регистрация приложения**

Чтобы использовать все возможности API ВКонтакте, необходимо было зарегистрировать приложение. После регистрации приложения были получены ID приложения (идентификатор приложения) и защищенный ключ. Они необходимы при составлении http-запросов к специальному серверу для получения информации из базы данных vk.com.

### **4.2. Авторизация пользователя**

Социальная сеть «ВКонтакте» обеспечивает пользователям или группам приватность некоторых их данных. Многое зависит от того, кто просматривает страницу: кто-то увидит на ней всю ту же информацию, что и владелец, а кто-то — лишь общедоступные данные.

В API этот принцип сохраняется. Почти все методы требуют авторизации пользователя перед началом работы. Проще говоря, сервер должен знать, кто к нему обращается за информацией, чтобы предоставить ее в том же виде, что и в основной версии сайта. Именно поэтому перед тем, как пользоваться приложением, которое рассматривается в данной работе, пользователю необходимо авторизоваться.

За авторизацию пользователя отвечает функция Authorization, которая с помощью элемента управления WebBrowser, отправляет специальный запрос (Листинг 1). После выполнения запроса откроется окно с запросом прав. В нем отображаются название приложения, иконки прав доступа, и имя, фамилия пользователя (рис. 1). После нажатия



на кнопку «Разрешить» пользователь будет авторизован.

```
1 private void Authorization()  
2 {  
3     string appid = "5648550";  
4     string scope = "groups,friends";  
5     string url = "https://oauth.vk.com/authorize?client_id=" + appid +  
6         "&display=popup&redirect_uri=https://oauth.vk.com/blank.html&scope=" +  
7         scope + "&response_type=token&v=5.56&state=123456";  
8     Auth authWindow = new Auth();  
9     authWindow.Show();  
10    WebBrowser browser = (WebBrowser)authWindow.FindName("WebBrowser");  
11    browser.Navigate(url);  
12 }
```

Листинг 1. Метод авторизации

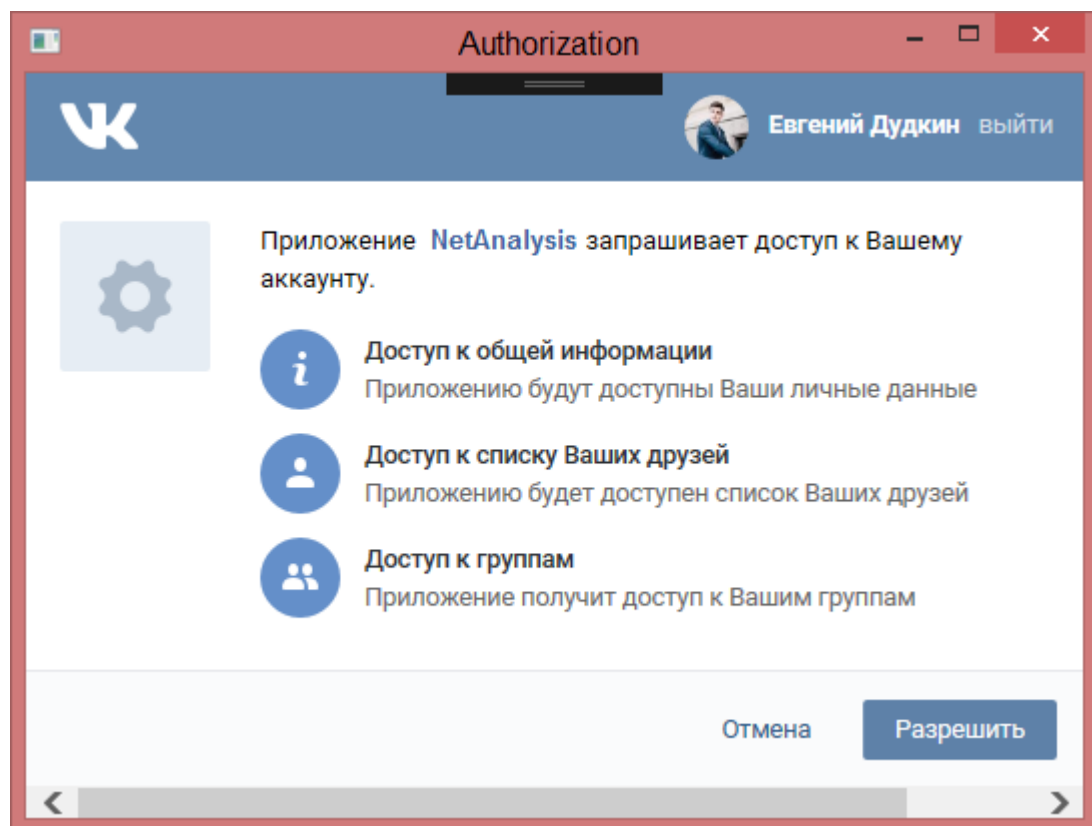


Рис. 1: Окно с запросом прав доступа

### 4.3. Поиск онлайн-сообществ

После успешной авторизации пользователя программа предоставляет возможность анализировать как связи между сообществами, так и связи внутри сообщества. Для выбора необходимых групп в программе, рассматриваемой в данной работе, реализован поиск онлайн-сообществ.

Пользователь может искать необходимые группы по ключевому слову, как и в основной версии сайта «ВКонтакте». Для этого необходимо ввести нужное ключевое слово в соответствующем поле ввода и нажать на кнопку «Поиск». После нажатия на кнопку вызывается метод, который запрашивает у сервера группы, в названии которых встречается это ключевое слово. Затем полученные группы отображаются на форме. На экран выводятся названия, идентификаторы и аватарки сообществ. Также напротив каждой группы отображается кнопка, позволяющая добавить сообщество в набор. В случае, если сообщество уже содержится в этом наборе, то нажатие на кнопку будет удалять сообщество из этого списка. На рисунке 2 представлен пример работы поиска по ключевому слову.

Также пользователь может искать группы по их идентификаторам. Для этого в соответствующем поле ввода необходимо ввести через запятую идентификаторы сообществ и нажать на кнопку «Поиск». После нажатия на кнопку на сервер отправится запрос сообществ по этим идентификаторам. Если такие идентификаторы действительно существуют, то соответствующие им группы отобразятся на экране. В противном случае появится сообщение о том, что поиск не дал результатов. Пример работы поиска по идентификаторам можно увидеть на рисунке 3.

Кроме того пользователь может воспользоваться своими сообществами. После нажатия на кнопку «Мои группы», на экране отобразятся все сообщества, в которых состоит пользователь.

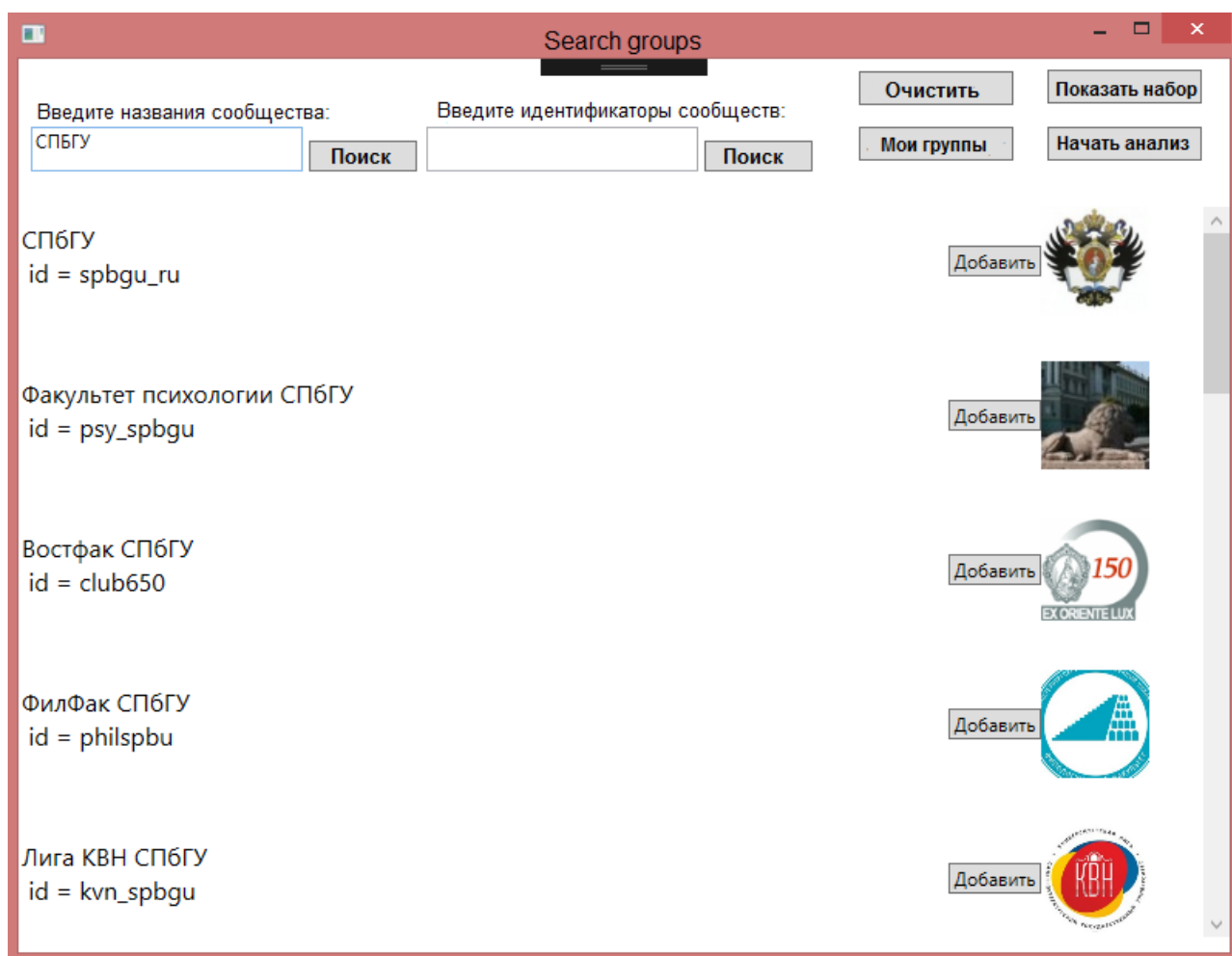


Рис. 2: Поиск по ключевому слову

Также на форме расположены некоторые вспомогательные кнопки. С помощью кнопки «Показать набор» можно узнать какие сообщества были добавлены пользователем в набор для анализа. Пример работы данной функции можно увидеть на рисунке 4. Также есть кнопка, позволяющая удалить все группы из этого набора.

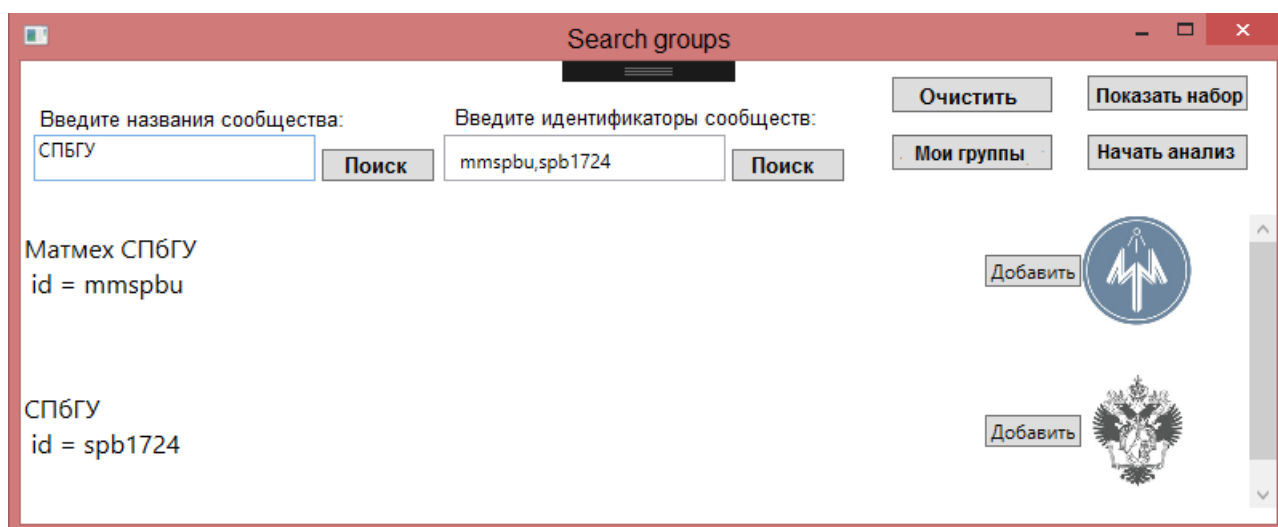


Рис. 3: Поиск по идентификатору

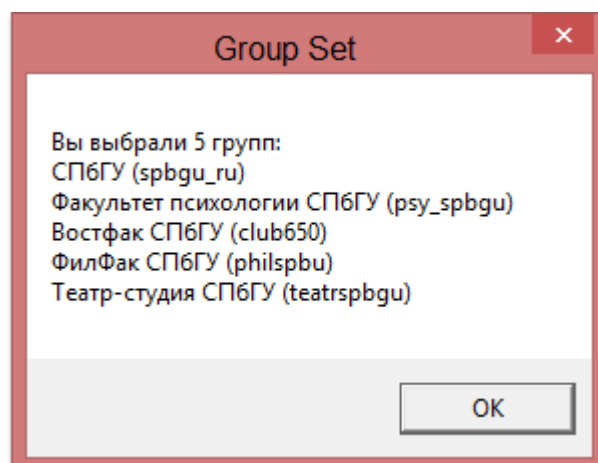


Рис. 4: Пример выбранных групп

#### 4.4. Построение социальной сетевой схемы сообщества

Если пользователь выбрал сообщество для анализа, то при нажатии кнопки «Начать анализ» программа предоставит ему различные типы связей. Связи нужны для того, чтобы формировать ребра между вершинами графа. Вершинами графа в социальной сетевой схеме являются участники сообщества. А ребро между вершинами будет существовать только тогда, когда между участниками в социальной сети есть соответствующая связь. Программа предоставляет следующие типы связей: город, родной город, страна, школа, вуз, дружба и наличие

общих сообществ.

На форме пользователь может выбрать необходимые ему связи и нажать на кнопку «Старт». После нажатия на кнопку вызовется функция `MainFunction`, которая и отвечает за создание графа. Для начала, с помощью `http`-запроса получается список участников сообщества. Это и будут вершины нашего графа. Затем для каждой выбранной связи, тоже с помощью запросов, получаются нужные данные об участниках. На основании этих данных получается список ребер, соответствующий этой связи. После этого ребра добавляются граф и помечаются своим цветом.

Так как для анализа данных был выбран язык `R`, а именно его библиотека `Igraph`, то граф будем формировать, как набор вершин и набор ребер. С помощью метода `make_empty_graph` создается пустая структура, которая в дальнейшей будет заполняться. После того, как получается список участников сообщества, в структуру добавляются имена этих участников, как вершины графа (Листинг 2). Список ребер соответствующей связи также добавляется в структуру. Реализацию формирования ребер и добавления их в граф, на примере связи по городу, можно увидеть на Листинге 3.

```
1 StringBuilder vert = new StringBuilder("");
2 engine.Evaluate("gr <- make_empty_graph(directed = FALSE)");
3 for (int i = 0; i < listOfUsersInfo.Count; i++)
4 {
5     vert.Append("'" + listOfUsersInfo[i].id.ToString() + " " +
6         listOfUsersInfo[i].last_name.ToString() + " " +
7         listOfUsersInfo[i].first_name.ToString() + "',");
8 }
9 vert.Remove(vert.Length - 1, 1);
10 engine.Evaluate("gr <- gr + vertices(" + vert.ToString() + ")");
```

Листинг 2. Добавление вершин

```

1  StringBuilder vect = new StringBuilder("");
2  if (cityFlag)
3  {
4      for (int i = 0; i < listOfUsersInfo.Count; i++)
5      {
6          Users tempUser = listOfUsersInfo[i];
7          if (tempUser.city == null)
8              continue;
9          for (int j = i + 1; j < listOfUsersInfo.Count; j++)
10         {
11             Users anotherUser = listOfUsersInfo[j];
12             if (anotherUser.city == null)
13                 continue;
14             if (tempUser.city.id == anotherUser.city.id)
15             {
16                 vect.Append("'" + listOfUsersInfo[i].id.ToString() + " " +
17                     listOfUsersInfo[i].last_name.ToString() + " " +
18                     listOfUsersInfo[i].first_name.ToString() + "', '" +
19                     listOfUsersInfo[j].id.ToString() + " " +
20                     listOfUsersInfo[j].last_name.ToString() + " " +
21                     listOfUsersInfo[j].first_name.ToString() + "',");
22             }
23         }
24     }
25     if (vect.Length != 0)
26     {
27         vect.Remove(vect.Length - 1, 1);
28         vectString = vect.ToString();
29         engine.Evaluate("gr <- add_edges(gr,c(" + vectString + ")," +
30             "color = \"green\")");
31     }
32 }

```

Листинг 3. Добавление ребер

## 4.5. Построение социальной сетевой схемы набора сообществ

После того, как составлен набор сообществ, который необходимо проанализировать, программа предоставляет пользователю различные типы связей. Связи здесь также нужны для того, чтобы формировать ребра между вершинами графа. Но в данном случае вершинами графа в социальной сетевой схеме являются сами сообщества. А ребро между вершинами будет существовать только тогда, когда между сообществами есть соответствующая связь. Программа предоставляет следующие типы связей: город, ссылки, страна, контакты и наличие общих участников.

После выбора необходимых связей и нажатия на кнопку «Старт» вызывается функция `MainFunction`. Эта функция отвечает за создание графа. Вершинами графа будет набор сообществ. Для каждой выбранной связи, с помощью запросов к серверу «ВКонтакте», получаются необходимые данные о сообществах. На основании этих данных получается список ребер, соответствующий этой связи. После этого ребра добавляются граф и помечаются своим цветом.

Как и в случае с графом участников сообщества, граф представляется как набор вершин и набор ребер. С помощью метода `make_empty_graph` создается пустая структура, которая в дальнейшем будет заполняться. На Листинге 4 представлено добавление в структуру вершин графа. После этого в структуру добавляются ребра, полученные на основании выбранных пользователем связей.

```
1 engine.Evaluate("gr <- make_empty_graph(directed = FALSE)");
2 for (int i = 0; i < listGroup.Count; i++)
3 {
4     st.Append("'" + listGroup[i].id.ToString() +
5     " " + listGroup[i].name.ToString() + "',");
6 }
7 st.Remove(st.Length - 1, 1);
8 engine.Evaluate("gr <- gr + vertices(" + st.ToString() + ")");
```

Листинг 4. Добавление вершин

## 4.6. Визуализация

После обработки данных и представления их в виде графа программа предоставляет пользователю набор инструментов для работы с ним. Одной из возможностей является его отображение на экране. Программное обеспечение дает возможность визуализировать граф как статическое изображение с параметрами (размер и цвет вершин, размер и цвет названия вершины, цвет ребер, алгоритм отрисовки, наличие названия вершин) по умолчанию. На рисунке 5 представлена сетевая схема одного из сообществ в социальной сети «ВКонтакте» с параметрами по умолчанию. Данная схема представлена в виде графа со связями: дружба и обучение в одном ВУЗе. Также программа дает возможность пользователю настроить параметры самостоятельно. На рисунке 6 изображена схема того же сообщества, но уже измененная (был изменен алгоритм отрисовки, цвет вершин и наличие названия вершины). Кроме того, программное обеспечение позволяет визуализировать граф на интерактивной форме, где можно настраивать параметры и перемещать вершины графа. Внешний вид формы представлен на рисунке 7. На форме можно вручную менять цвета и размеры вершин, цвета и размеры ребер, алгоритмы отрисовки, наличие названия вершин.

Для больших графов предусмотрен специальный режим, который позволяет наглядно представить их структуру. На рисунке 8 представлен граф, состоящий из 550 вершин. При большом количестве изолированных вершин, можно воспользоваться отображением сетевой схемы без изолянтов. На рисунке 9 представлен граф, у которого доля изолированных вершин выше 95 %. Рисунок 10 иллюстрирует эту же структуру, но уже без изолянтов.



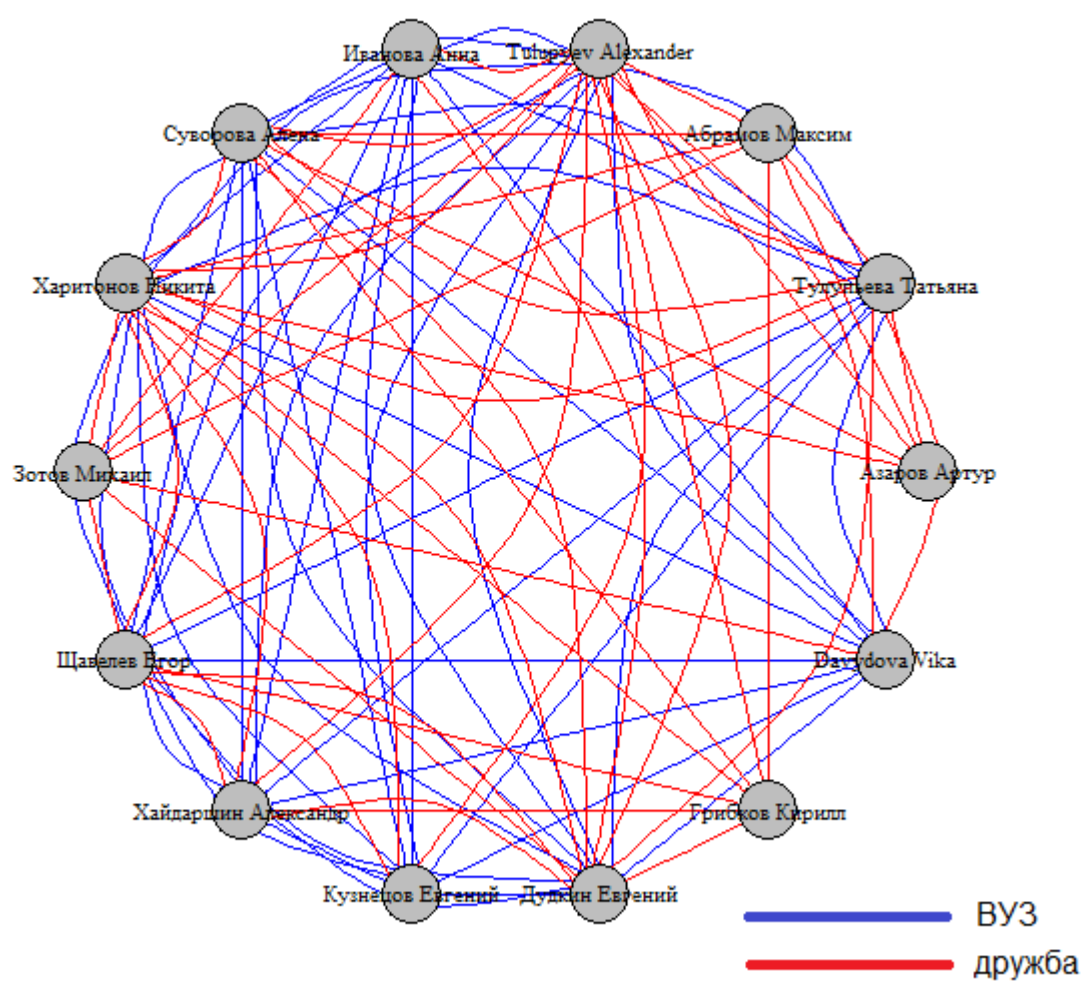


Рис. 5: Пример графа

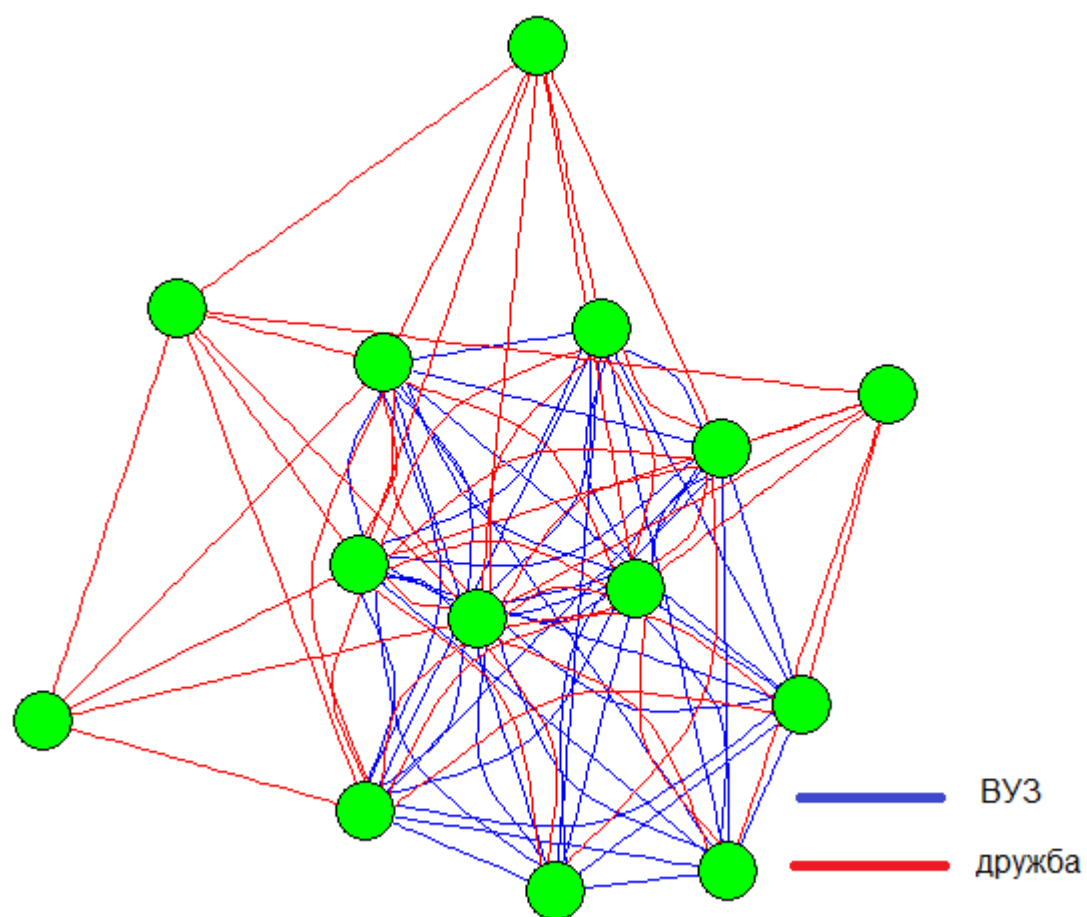


Рис. 6: Пример графа

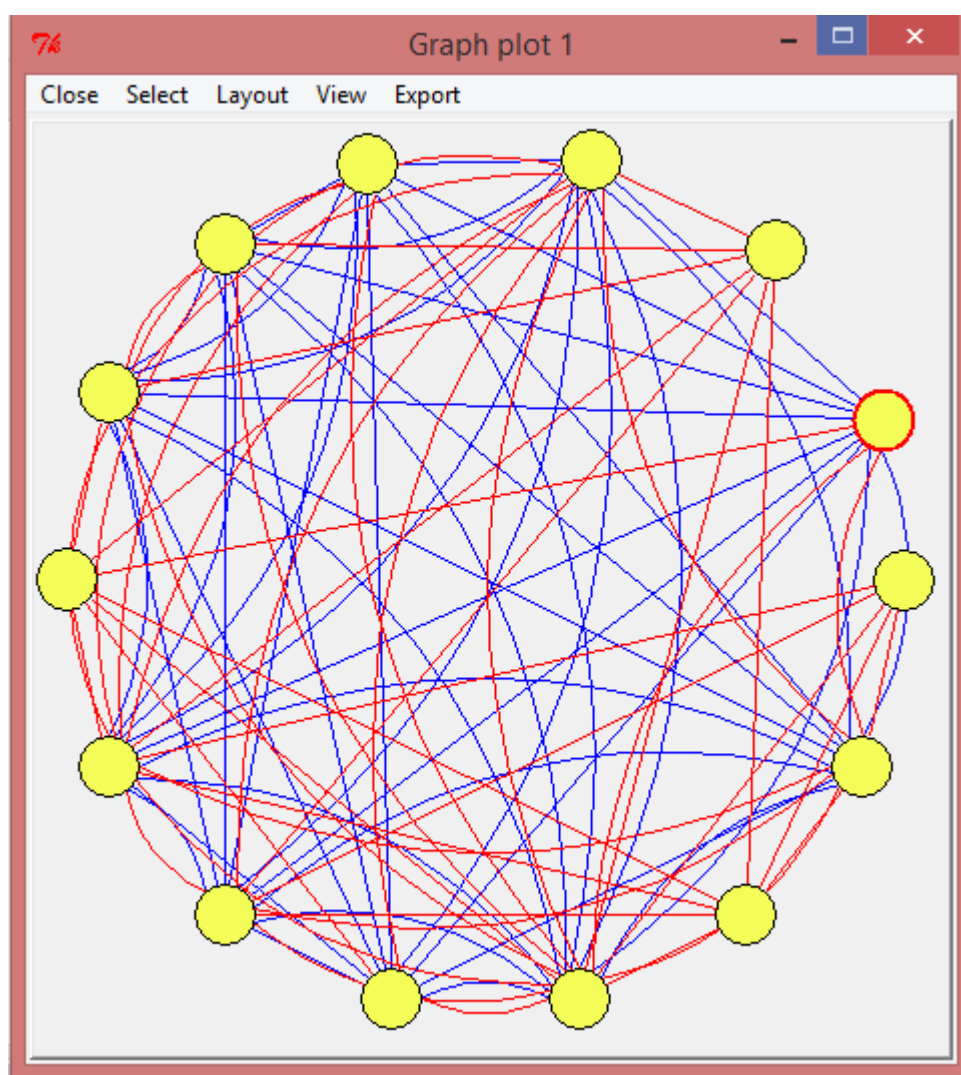


Рис. 7: Интерактивная форма

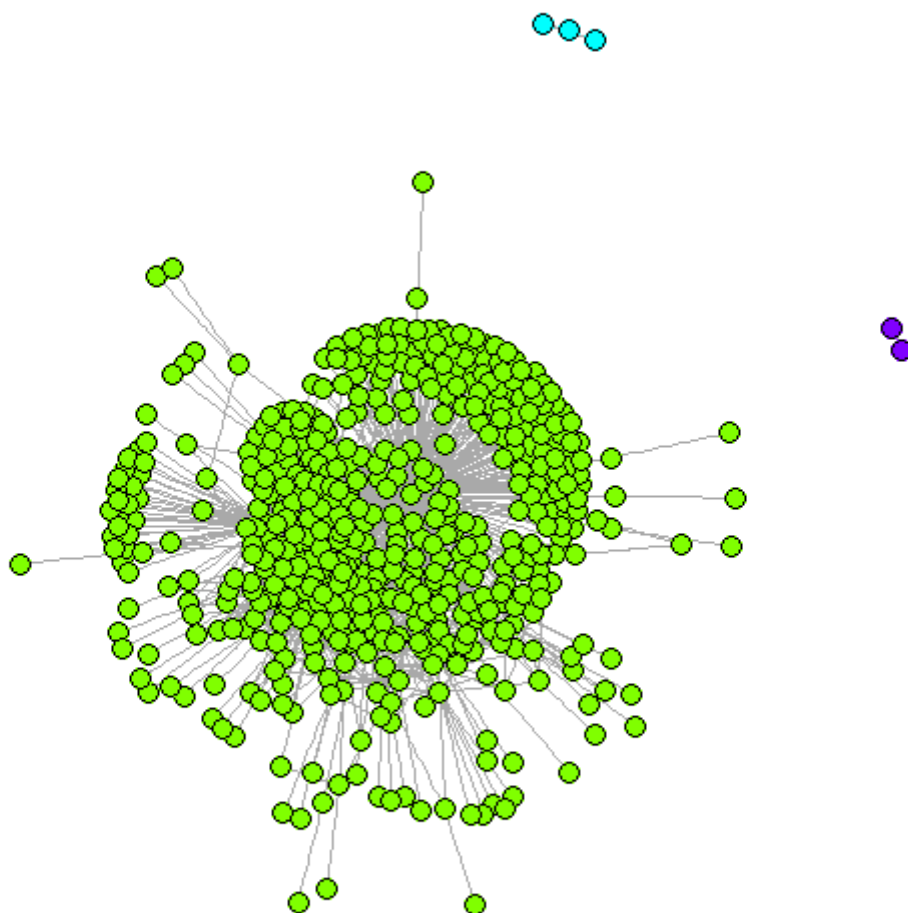


Рис. 8: Пример большого графа

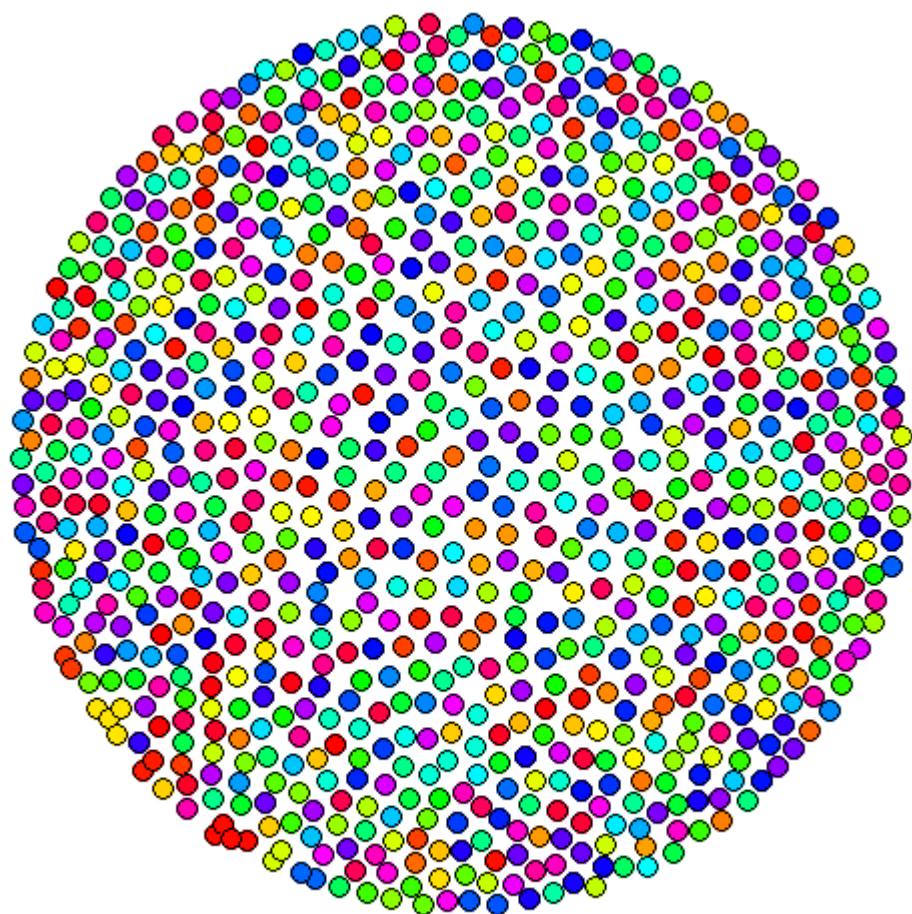


Рис. 9: Пример графа с высокой долей изолянтов

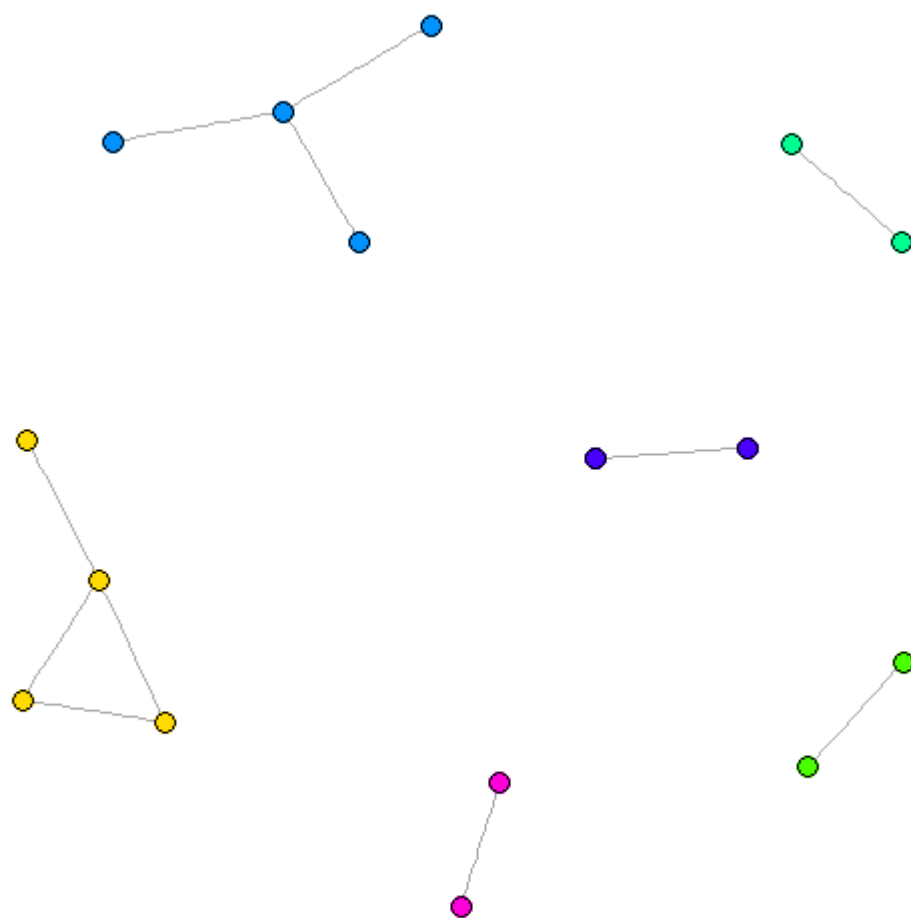


Рис. 10: Граф без изолянтов

## 4.7. Создание CSV файла

У пользователя должна быть возможность воспользоваться другими средствами визуализации или анализа. Такие программы принимают на вход файлы формата csv. Поэтому возникла необходимость представления графа в виде csv файла.

Граф онлайн-сообщества представляется как таблица из трех столбцов: Source, Target, Type. С помощью этой таблицы описываются все ребра графа. В столбцах Source и Target находятся имена и фамилии участников группы. То есть, если два участника группы являются друзьями в социальной сети и была выбрана связь «дружба», то при формировании графа появится ребро, которое будет соединять вершины с именами и фамилиями этих участников. А при формировании csv файла имя и фамилия одного участника появится в столбце Source, а другого - напротив, в столбце Target. Столбец Type во всех строках будет иметь значение «Undirected», т.к. граф является неориентированным.

При нажатии на кнопку «Создать CSV файл» вызывается метод MakeCSVButtonClick, который получает список всех ребер графа и представляет их в виде трех столбцов, как было описано выше (Листинг 5).

```
1 private void MakeCSVButtonClick(object sender, RoutedEventArgs e)
2 {
3     CharacterMatrix v = engine.Evaluate("v <- as_edgelist(gr)").AsCharacterMatrix();
4     StringBuilder csv = new StringBuilder();
5     csv.AppendLine("Source,Target,Type");
6     for (int i = 0; i < v.RowCount; i++)
7     {
8         csv.AppendLine(v[i, 0].ToString() + "," + v[i, 1].ToString()+ ",Undirected");
9     }
10    SaveFileDialog dlg = new SaveFileDialog();
11    dlg.FileName = groupName ; // Default file name
12    dlg.DefaultExt = ".csv"; // Default file extension
13    dlg.Filter = "Text documents (.csv)|*.csv"; // Filter files by extension
14
15    if (dlg.ShowDialog() == true)
16        File.AppendAllText(dlg.FileName, csv.ToString());
17 }
```

Листинг 5. Создание csv файла

После нажатия на кнопку появляется окно, которое позволяет выбрать место сохранения файла и его название (рис. 11).

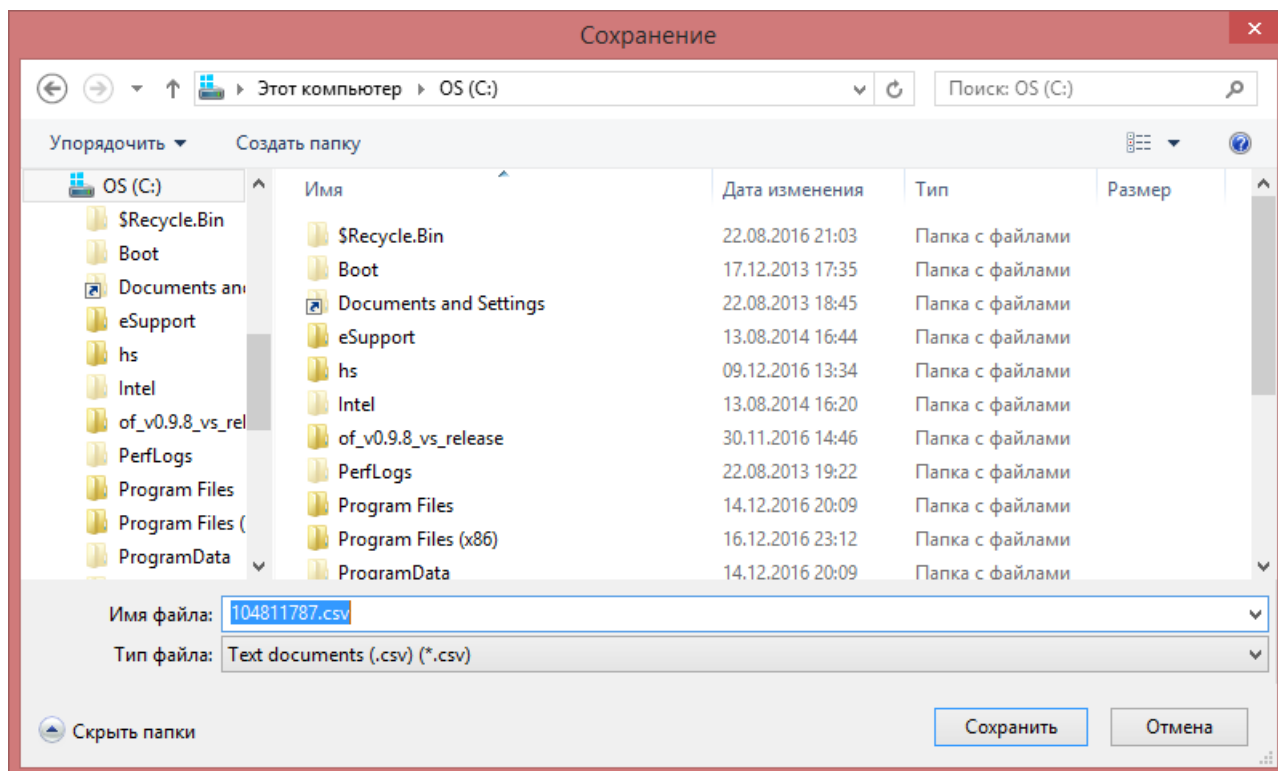


Рис. 11: Окно сохранения

После сохранения файла можно воспользоваться другими средствами визуализации. Например, программным обеспечением Gephi [1] для визуализации всех видов графиков и сетей. Так для сетевой схемы онлайн-сообщества, представленного на рисунке 5, был создан csv файл. Затем этот файл был дан на вход программе Gephi. Результат визуализации средствами программного обеспечения Gephi можно видеть на рисунке 12.

Кроме того, программное обеспечение, рассматриваемое в данной работе, позволяет импортировать данные из csv файлов. Таким образом, можно строить, визуализировать и анализировать графы, представленных в виде csv файлов (состоящие из столбцов Source, Target и Type). Данная функция позволяет использовать уже имеющиеся данные об онлайн-сообществе, повторно не обрабатывая их.



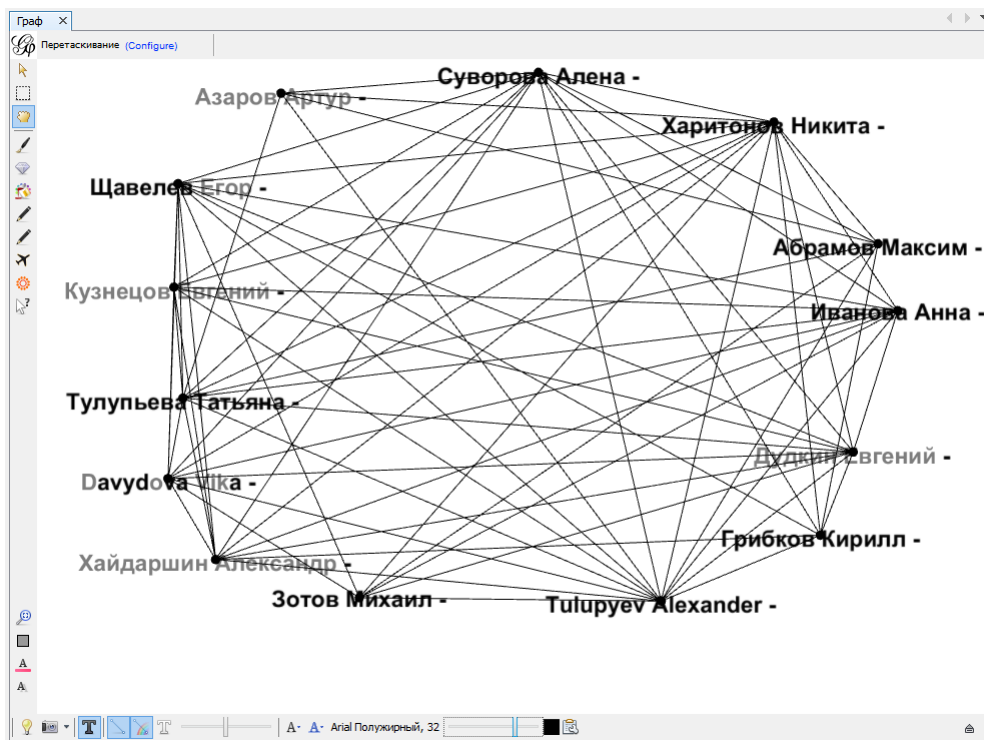


Рис. 12: Граф в программе Gephi

## 4.8. Анализ графа

Программное обеспечение, рассматриваемое в данной работе, позволяет пользователю считать некоторые базовые метрики графа:

- количество вершин;
- число ребер;
- плотность;
- количество изолированных вершин (их доля в процентах относительно общего количества вершин);
- количество компонент связности;
- количество вершин в каждой компоненте связности;
- доля вершин максимальной компоненты связности (в
- средняя геодезическая дистанция;

- центральность по посредничеству каждого узла;
- средняя центральность по посредничеству (без изолянтов);
- центральность по близости каждого узла;
- средняя центральность по близости (без изолянтов);
- центральность собственного вектора каждого узла;
- средняя центральность собственного вектора (без изолянтов);
- центральность по степени каждого узла;
- средняя центральность по степени (без изолянтов);
- взаимность;
- транзитивность .

## 4.9. Пример использования

Рассмотрим один из примеров реального исследования, для которого потенциально необходимо разработанное программное обеспечение. В 2016 году была опубликована статья «Структура и функции онлайн-сообществ: сетевая картография ВИЧ-релевантных групп в социальной сети «ВКонтакте» » [17]. В исследовании поднимается вопрос: можно ли проследить соответствие между целями и структурой онлайн-сообществ, а именно групп «ВКонтакте», объединенных темой ВИЧ/СПИД [17].

Объектом исследования были онлайн-сообщества социальной сети «ВКонтакте», связанные с тематикой ВИЧ/СПИД. Из 987 групп, релевантных данной теме, была составлена итоговая выборка, которая содержала 15 сопоставимых по численности онлайн-сообществ. Социальная структура группы была представлена в виде графа, вершинами которого являются ее участники, а ребро между двумя вершинами существует тогда, когда участники являются друзьями в рамках социальной сети. Для каждой группы из выборки была определена ее тема и

цель. Под целью группы подразумевалось ее наблюдаемое предназначение и желаемый результат [17]. Определялись цели групп, исходя из ее названия, информации, указанной в разделе «описание», и содержания последних постов на «стене».

После этого была проведена содержательная классификация онлайн-групп. В исследуемой совокупности были выделены основные их типы в зависимости от их тематики и целей: I. Группы общественных движений: 1.1. ВИЧ-активисты (борьба с распространением вируса и дискриминацией ВИЧ-инфицированных, просвещение и информирование населения о ВИЧ, борьба против СПИД-диссидентов) (5 групп). 1.2. СПИД-диссиденты (отрицающие существование ВИЧ и/или связь между ВИЧ и СПИД) (2 группы). II. Группы для знакомства и общения ВИЧ-инфицированных пользователей (4 группы). III. Группы социальной (информационной, психологической) поддержки больных, объединяющие равнодушных пользователей (2 группы). IV. Группы, представляющие в онлайн-среде офлайн-организации и проекты (СПИД-центры, фонды, больницы и др.) (2 группы) [17]. По мнению исследователей, данные типы отражают цели сообществ.

После этого был проведен анализ структуры каждой группы из итоговой выборки. На основании этого анализа было предложено различать пять типов структуры онлайн сообществ: консолидированная толпа (одно крупное и плотное ядро), биполярная толпа (два кластера), стратифицированная структура (одно крупное и плотное ядро; сравнительно высокая централизация графа), кластерная структура (сеть состоит из нескольких плотных кластеров, слабо связанных между собой; мало изолянтов), дезинтегрированная (подавляющая доля изолянтов, крайне низкая плотность, отсутствие сообщества как такового), неопределенная (противоречивая интерпретация) [17]. После этого были установлены соответствия типов структур с типами целей и тем.

Анализ структуры производился следующим образом: для каждой группы из выборки строилась ее сетевая схема (граф, вершины которого - участники сообщества, а ребро между вершинами существует тогда, когда участники являются друзьями в рамках социальной сети), по по-

лученному графу считались структурные характеристики сообщества, происходила визуализация графа. В качестве характеристик использовались следующие метрики: общая численность группы, доля изолянтов в процентах, число ребер, плотность (без изолянтов), доля вершин максимальной компоненты в процентах, средняя степень центральности (без изолянтов), средняя геодезическая дистанция, централизация графа по степени (без изолянтов).

Этап анализа структуры каждой группы из выборки можно реализовать с помощью программного обеспечения, рассматриваемого в данной работе. Для подтверждения этого была составлена тестовая выборка из семи сообществ релевантных тематике ВИЧ/СПИД. После этого была составлена сетевая схема каждой группы из этой выборки. Затем считались те же структурные характеристики сообщества, что и в исследовании, описанном выше (полученные результаты представлены в Таблице 1). После этого каждая сетевая схема была визуализирована средствами программного обеспечения.

Исследование, рассмотренное выше, показало, что онлайн-сообщества используются для разных целей, а цели связаны с сетевой схемой сообщества [17]. В рамках исследования были выделены пять структурных паттернов, установлены соответствия паттернов с типами целей и тем. Тип консолидированная толпа представляет собой сетевую схему, состоящую из одного большого ядра пользователей и нескольких малых компонент связности. Данный паттерн характерен для тех групп, которые предназначены для знакомств ВИЧ-инфицированных людей. В выборке, описанной в данной работе, содержатся два сообщества знакомств ВИЧ-инфицированных людей. Их сетевые схемы представлены на рисунках 13 и 14.

Таблица 1: Структурные характеристики онлайн-сообществ

Описание группы	Общая численность группы	Доля изолянтов, в %	Число ребер	Доля вершин макс.комп. (в %)	Плотность (без изолянтов)	Средняя степень центр. (без изол.)	Средняя геодезич. дист.	Централизация графа по по степ. (без изол.)
Стоп СПИД	808	34	1458	65	0,01	5,47	2,4	0,73
ВИЧ+ знакомства	1828	57	1423	40	0,005	3,63	3,89	0,10
СПИД центр	741	51	744	39	0,011	4,06	3,7	0,17
ВИЧ+ Украина	770	58	653	38	0,012	4,04	5,09	0,12
Против СПИДА	897	98	11	0,4	0,09	1,37	1,31	0,10
ВИЧ диссиденты	1059	53	1865	46	0,015	7,46	3,1	0,34
Поддержка людей с ВИЧ	668	81	114	5	0,015	1,8	2,64	0,082

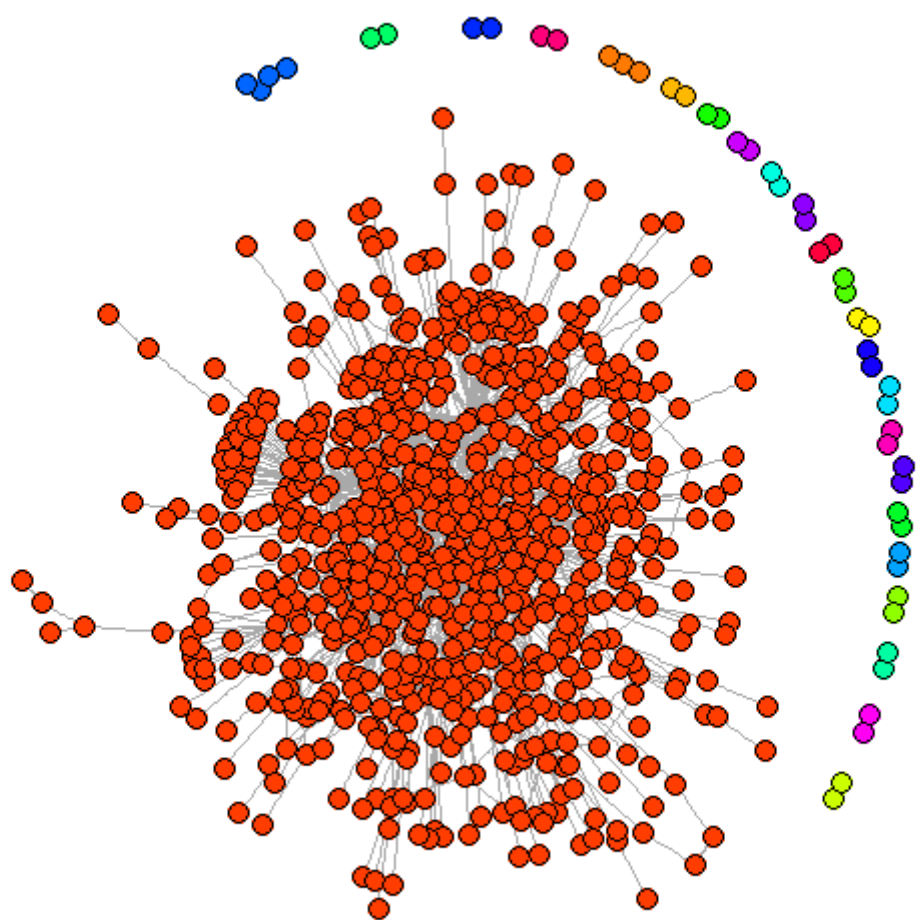


Рис. 13: ВИЧ+ знакомства

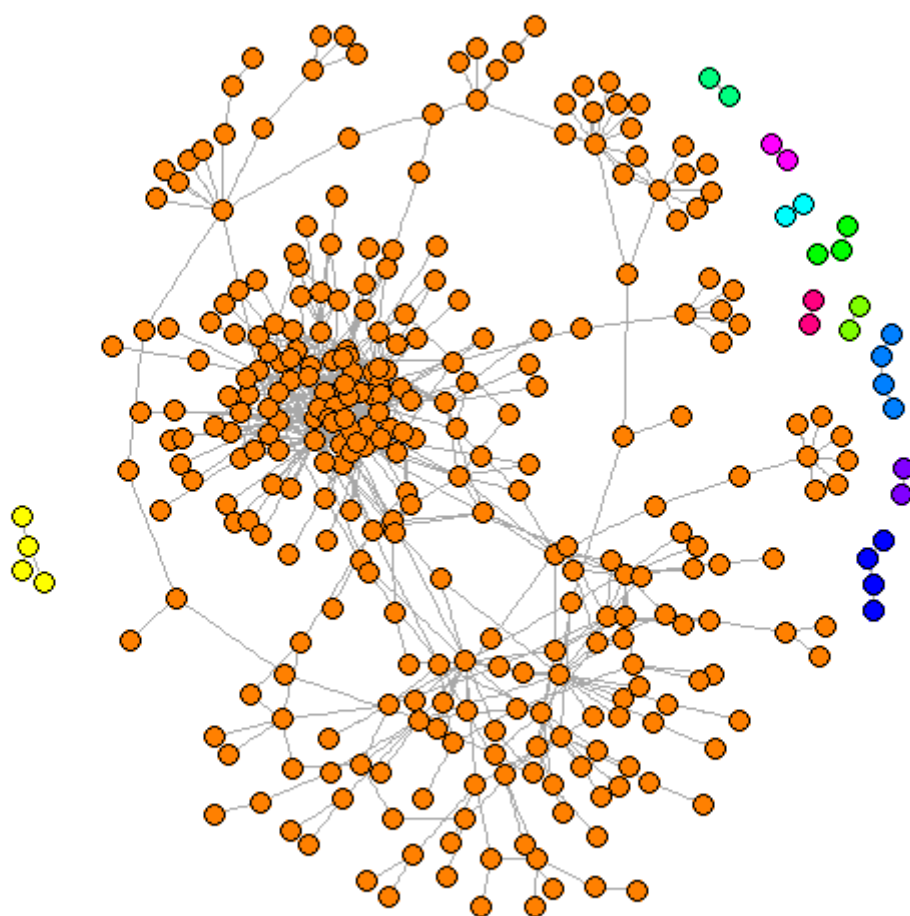


Рис. 14: ВИЧ+ Украина

Тип кластерная структура представляет схему с несколькими компонентами связности. В такой схеме отсутствует какое-либо большое ядро пользователей. Такая структура, по мнению авторов статьи, характерна для групп активистов. В наборе имеется один представитель таких групп. Его сетевая схема представлена на рисунке 15.

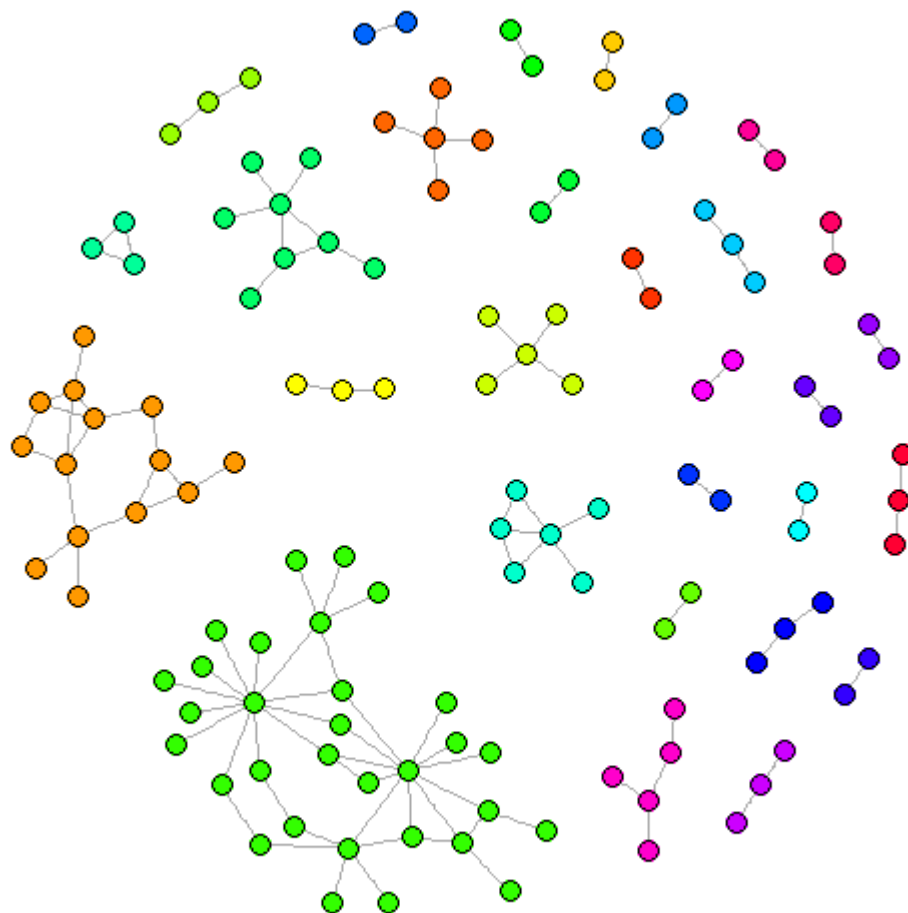


Рис. 15: Поддержка людей с ВИЧ

Тип дезинтегрированная структура характеризуется доминированием изолянтов в сети (более 90%). В исследовании такой структурный паттерн соответствует группам против СПИД. Рисунок 16 иллюстрирует сетевую схему такой группы из выборки, составленной в рамках данной работы. Что же касается стратифицированной структуры, то этот структурный паттерн характеризуется одним крупным и плотным ядром и сравнительно высокой централизацией графа. В результатах исследования говорится, что стратифицированная структура соответствует группам ВИЧ-активистов и, особенно, группам ВИЧ-диссидентов. Сетевая схема группы ВИЧ-диссидентов из тестовой выборки представлена на рисунке 17.



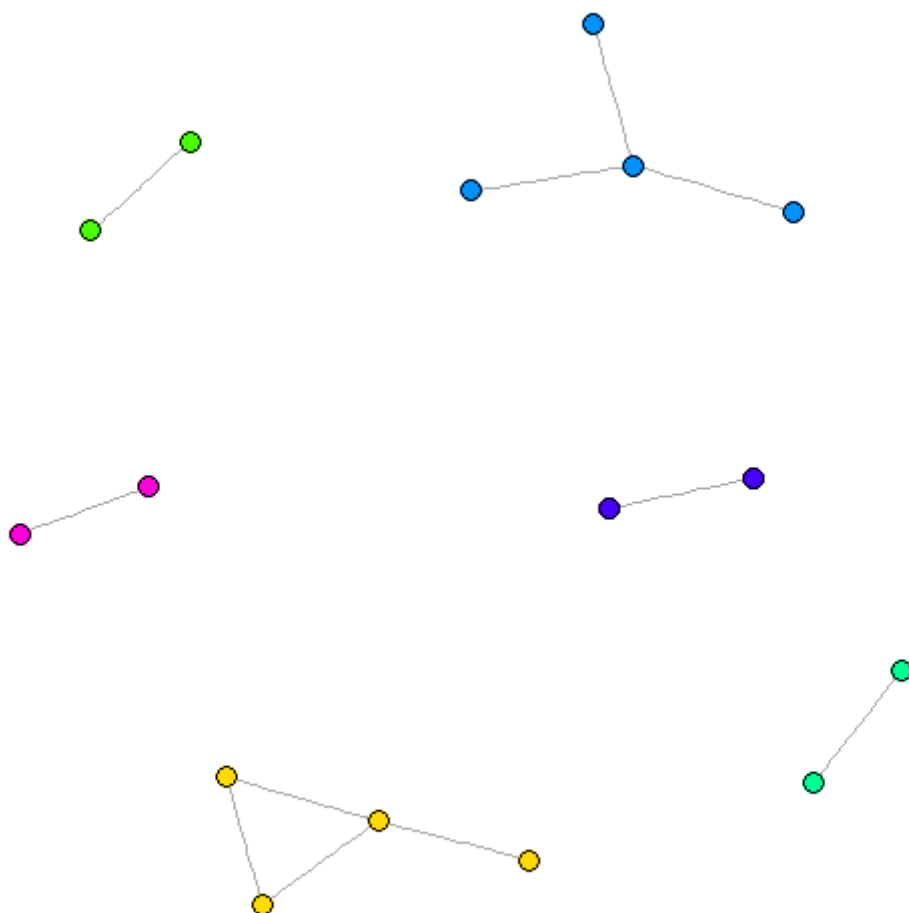


Рис. 16: Против СПИДА

Таким образом, программное обеспечение, рассматриваемое в данной работе, способно решать задачи, которые стоят перед исследователями, занимающимися анализом структуры онлайн-сообществ. Этот вывод следует из того, что набор инструментов, предоставляемый программным обеспечением, решает все технические задачи приведенного примера исследования. Этап построения, анализа и визуализации сетевых схем ВИЧ-релевантных онлайн-сообществ можно без труда реализовать с помощью данного программного обеспечения. Также необхо-

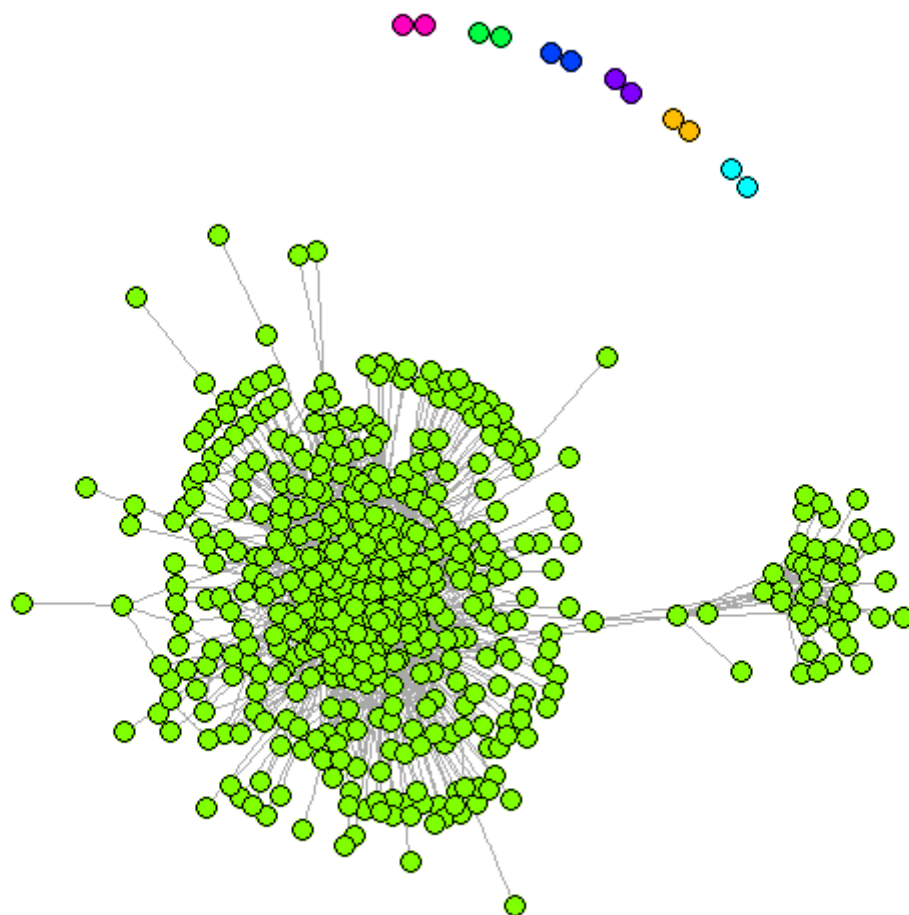


Рис. 17: ВИЧ диссиденты

димо отметить, что программное обеспечение помимо дружеской связи участников сообщества, рассмотренной в статье, предлагает еще и другие связи, которые могут быть интересны для научных деятелей. Кроме того, с помощью программного обеспечения можно находить связи между анализируемыми онлайн-сообществами. Например, на итоговой выборке из семи сообществ были выявлены следующие связи: страна (рис. 18), наличие общих участников (рис. 19), город (рис. 20).

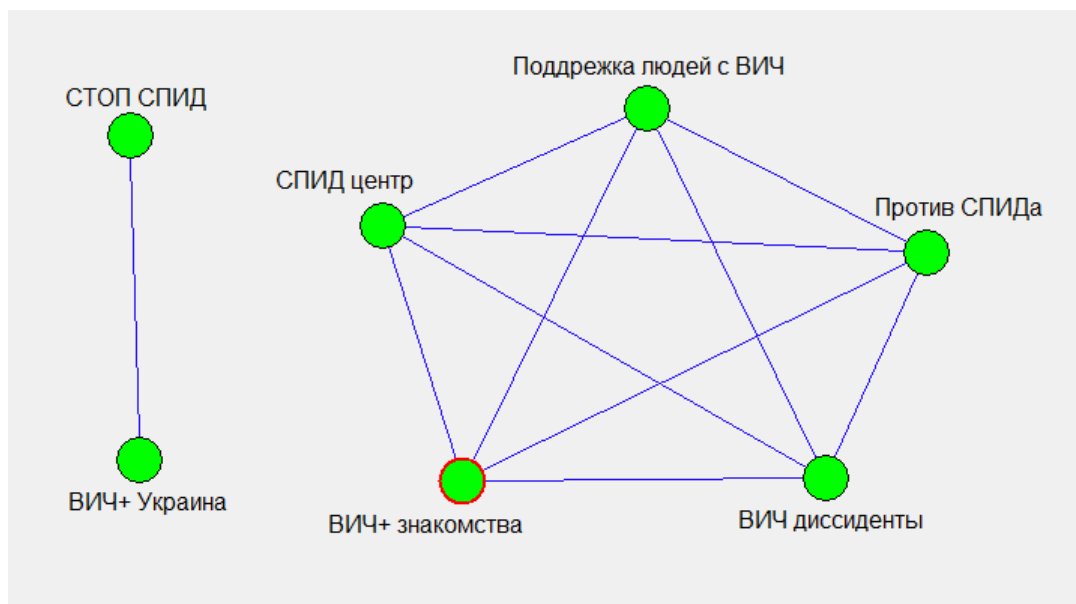


Рис. 18: Сетевая схема со связью «страна»

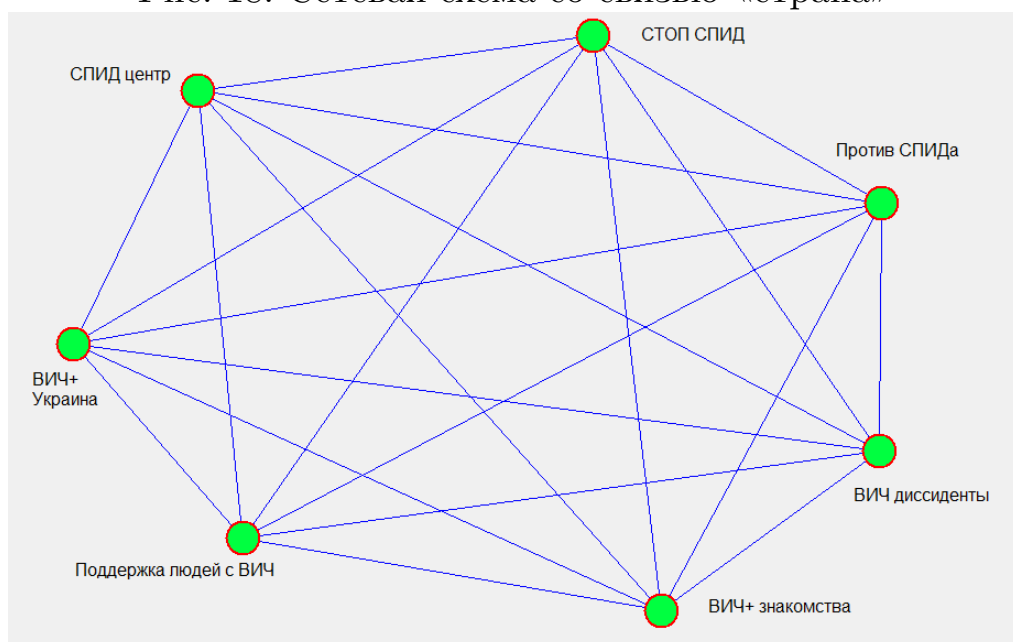


Рис. 19: Сетевая схема со связью «наличие общих участников»

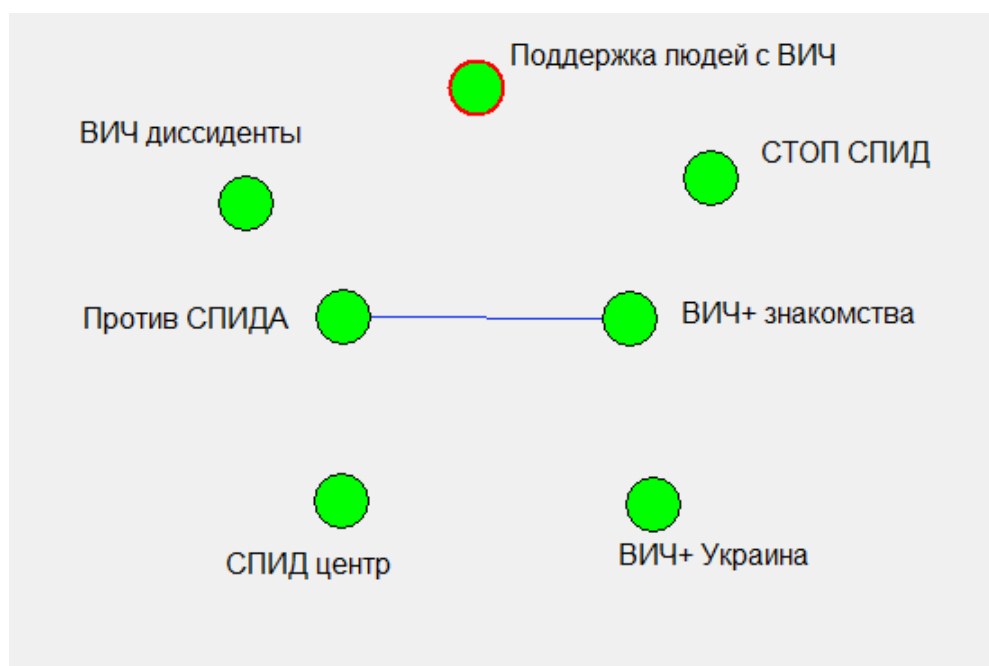


Рис. 20: Сетевая схема со связью «город»

## Заключение

Таким образом были решены все поставленные задачи. Был обеспечен сбор данных о структурах онлайн-сообществ социальной сети «ВКонтакте», включая список контактов, город, ссылки, страна и информацию об участниках. Была обеспечена визуализация структуры сообщества в виде графа, вершинами которого являются участники этого сообщества, а ребрами — соответствующие связи между этими пользователями (дружба, обучение в одном вузе, город, родной город, обучение в одной школе, наличие общих сообществ). Была обеспечена визуализация структуры связей между несколькими сообществами в виде графа, вершинами которого являются эти сообщества, а ребрами — соответствующие связи между ними (город, ссылки на одни и те же ресурсы или на друг друга, страна, наличие общих пользователей в графе контакты). Были предоставлены инструменты для базового анализа полученных графов (вычисление показателей центральности, среднего для показателей центральности, количества вершин и ребер, вычисление количества изолированных вершин, количества компонент связности, количества вершин максимальной компоненты связности и

др.). Была предоставлена возможность импортировать данные о структурах в виде CSV файлов.

Таким образом, цель — автоматизировать сбор, обработку и визуализацию информации об онлайн-сообществах в социальной сети «ВКонтакте», можно считать достигнутой.

Работа выполнялась в рамках исследований, проводимых в лаборатории теоретических и междисциплинарных проблем информатики СПИИРАН. Результаты работы были представлены на конференции СПИСОК-2017, доклад «Интер- и интрасетевые связи: визуализация и анализ».

## Список литературы

- [1] Gephi - The Open Graph Viz Platform. — URL: <http://gephi.org> (дата обращения: 9.11.2016).
- [2] Hummel H. G. K. et al. Encouraging contributions in learning networks using incentive mechanisms // Journal of computer assisted learning. — 2005. — Vol. 21, no. 5. — P. 355–365.
- [3] IQBuzz. — URL: <http://iqbuzz.pro/> (дата обращения: 5.05.2017).
- [4] Martinez M. G., Walton B. The wisdom of crowds: The potential of online communities as a tool for data analysis // Technovation. — 2014. — Vol. 34, no. 4. — P. 203–214.
- [5] Quercia D. et al. The personality of popular facebook users // Proceedings of the ACM 2012 conference on computer supported cooperative work. — 2012.
- [6] R.NET - Home. — URL: <http://rnet.codeplex.com> (дата обращения: 19.11.2016).
- [7] Ramo D. E., Prochaska J. J. Broad reach and targeted recruitment using Facebook for an online survey of young adult substance use // Journal of medical Internet research. — 2012. — Vol. 14, no. 1. — P. e28.
- [8] Ren Y. et al. Building Member Attachment in Online Communities: Applying Theories of Group Identity and Interpersonal Bonds // Mis Quarterly. — 2012. — Vol. 36, no. 3. — P. 841–864.
- [9] Smith M.A., Rainie L., Shneiderman B. Mapping twitter topic networks: From polarized crowds to community clusters // Pew Research Internet Project. — 2014.
- [10] Sun N., Rau P.P.L., Ma L. Understanding lurkers in online communities: A literature review // Computers in Human Behavior. — 2014. — Vol. 38. — P. 110–117.

- [11] Wasserman S., Faust K. Social networks analysis: Methods and applications. — New York: Cambridge University, 2005.
- [12] Wikipedia. — URL: [https://en.wikipedia.org/wiki/VK\\_\(social\\_networking\)](https://en.wikipedia.org/wiki/VK_(social_networking)) (дата обращения: 20.05.2017).
- [13] YouScan. — URL: <https://youscan.io/> (дата обращения: 20.05.2017).
- [14] igraph – Network analysis software. — URL: <http://igraph.org/> (дата обращения: 25.11.2016).
- [15] Разработчикам | ВКонтакте. — URL: <https://vk.com/dev> (дата обращения: 20.11.2016).
- [16] Рыков Ю.Г. Сетевое неравенство и структура онлайн-сообществ // Журнал социологии и социальной антропологии. — 2015. — Т. XVIII, № 4. — С. 144–156.
- [17] Рыков Ю. Г., Кольцова Е. Ю., Мейлахс П. А. Структура и функции онлайн-сообществ: сетевая картография ВИЧ-релевантных групп в социальной сети «ВКонтакте» // Социологические исследования. — 2016. — № 8. — С. 30–42.
- [18] Шерстобитов А. С., Брянов К. А. Технологии политической мобилизации в социальной сети «ВКонтакте»: сетевой анализ протестного и провластного сегментов // Исторические, философские, политические и юридические науки, культурология и искусствоведение. Вопросы теории и практики. — 2013. — № 10-1. — С. 36.